Journal of Scientific Research and Development 2 (12): 24-32, 2015
Available online at www.jsrad.org
ISSN 1115-7569

The evaluation of robust and efficient estimators for log-logistic distribution for censored data with/without outliers)

Ibrahim Arik, Yeliz Mert Kantar *, Ismail Yenilmez

Department of Statistics, Anadolu University, Eskisehir, Turkey

Abstract: In the case of samples containing censored values, the Cox regression commonly used under the assumption of proportional hazards is not always applicable. The accelerated failure time model as parametric model provides an alternative to the Cox model. Parametric regression models assume a distribution type model with unknown parameters to stand for times of failure of the interest. The log-logistic regression model, based on log-logistic distribution, is one of the alternative parametric models. In this study, we have studied the robust estimators of the log-logistic distribution in the cases of censored data with/without outliers. The classical and robust estimators of the parameters of the log-logistic distribution have been evaluated by means of Monte Carlo simulation study. The obtained results show that the certain robust estimators show good performance for the shape parameter of the log-logistic distribution in the cases of censored data with/without outliers.

Key words: Censored data; Censoring; Outlier; Parametric models; Robust estimators

1. Introduction

Survival analysis is concerned with data containing life times, waiting times, more generally times to the realization of the event of interest. Such data is generally expressed as survival data. Survival analysis is different from basic statistical methods in terms of containing censored data. (Cox and Oakes, 1984; Kleinbaum, 1996; Klein and Moeschberger, 1997; Lee and Wang, 2003).

In the literature, non-parametric approaches are dominant to examine censored data in survival analysis. Especially the Cox regression is commonly used tool. However, this regression model requires pre-conditions such as proportional hazards assumption. (Kleinbaum, 1996; Klein and Moeschberger, 1997; Lawless, 1982; Karasoy and Ata, 2008). Furthermore, it is well known that the Cox regression is not as efficient as other parametric models if parametric form is specified correctly (Basu et al., 2006). For these reasons, alternatively flexible parametric models based on statistical distributions such as Weibull, Gamma and Log-logistic, have been widely-used in literature and compared versus the Cox regression (May et. al 2004; Wei, 1992).

While working the parameter model in survival analysis, the following issues should be taken into consideration: For example, (i) the underlying distribution can be misspecified, (ii) maximum likelihood or other classical estimators may be severely affected by contamination of outliers and lead to very poor results (Basu et al., 2003), (iii) convergence problems can appear while working with estimators which need iterative numerical methods for obtaining estimates, (iv) we have only small data sets, or data sets with censored data.

The log-logistic distribution (LLD), a particular member of the Burr family, has non-negative random variables, is a useful distribution in survival analysis (Klein and Moeschberger, 1997; Marubini and Valsecchi, 1995; Pourhoseingholi et al., 2007; Gupta et al., 1999; Arik and Kantar, 2014) and it has non-monotonic hazard function which increases at early times and decreases at later times. Non-monotonic hazard function is convenient for modeling some sets of cancer survival data (Bennett, 1983). Gubta et al. (1999) have studied lung cancer data with LLD. They estimated the mortality ratio. Byers et al. (1988) have shown that LLD provided better fits than Weibull and logistic and Gompertz distribution for HIV data. Gruure (2015) have studied LLD with three real survival data.

Moreover, although LLD is similar to log-normal, it is more suitable for use in the survival data analysis (Bennett, 1983). Also, its cumulative distribution function is a simple explicit form. This allows us to get rid of computational complexity when fitting data with censoring since the survival function is obtained from cumulative distribution function for the censored observations.

In literature, the parameters of LLD for cases of censored and non-censored samples have been studied by a number of authors (Chen, 1997; Chen 2006; Kantam et al., 2008; Kantar and Arik, 2014). While Chen (1997, 2006) has considered the parameters of LLD for non-censored sample, Kantam et al. (2008) have studied the scale parameter of LLD

* Corresponding Author.

for type-II censored samples. On the other hand, Kantar and Arik (2014) have researched the performance of certain M-estimators to estimate the parameters of LLD for samples with/without outliers.

As stated by (Basu et al., 2003), it can be expressed that detection of contamination of outliers can be very difficult in the presence of censored data. In such cases, robust methods, which automatically decrease the effects of contamination, can serve to provide more reasonable results than classic estimators.

In this paper, we study robust estimators of LLD for censored data with and without contamination of outliers. Our aim is to find robust and efficient estimators among the considered estimators: maximum likelihood (ML), least squares (LS), M-estimator (Tukey), repeated median (RM), least trimmed squares (LTS) and least median square (LMS).

The organization of the paper is as follows: The Section 2 mentions LLD and its some distributional properties. Estimators used in study are introduced in Section 3. Simulation and its results are presented in Section 4. Finally Section 5 contains the conclusions of the study.

2. Log-logistic distribution

The probability density function, cumulative distribution function and quantile function of log-logistic (LL) random variable are respectively given as follows:

$$f(x) = \frac{\alpha}{\gamma}\left[\frac{x}{\gamma}\right]^{\alpha-1}\left[1+\left[\frac{x}{\gamma}\right]^{\alpha}\right]^{-2}, x \geq 0;\ \alpha,\gamma > 0 \quad (1)$$

$$F(x) = 1 - \left[1+\left[\frac{x}{\gamma}\right]^{\alpha}\right]^{-1}, \ x \geq 0 \text{ ve } \alpha,\gamma > 0 \quad (2)$$

$$Q(p) = \alpha\left[\frac{p}{1-p}\right]^{\frac{1}{\gamma}}, \ 0 \leq p \leq 1 \quad (3)$$

The first and second population moments are only defined when α>1 and α>2, respectively. Therefore moment estimators are not widely used. The estimation methods for parameters of LLD have been widely discussed in the literature (See, for examples: Fitzgerald, 1996; Ashkar and Mahdi, 2003; Ashkar and Mahdi, 2006; Singh and Guo, 1995; Chen, 2006; Kantam et al., 2008). Particularly, the estimation of shape parameter of the LLD is of great interest because it characterizes hazard rate. When α > 1, the hazard function becomes unimodal and when α ≤ 1, the hazard decreases monotonically (Guure, 2015).

Survival function and hazard function corresponding to LLD are respectively given as follows:

$$S(x) = \left[1+\left[\frac{x}{\gamma}\right]^{\alpha}\right]^{-1}, \ x \geq 0 \text{ ve } \alpha,\gamma > 0 \quad (4)$$

$$h(x) = \frac{\alpha}{\gamma}\left[\frac{x}{\gamma}\right]^{\alpha-1}\left[1+\left[\frac{x}{\gamma}\right]^{\alpha}\right]^{-1}, \ x \geq 0;\ \alpha,\gamma > 0 \quad (5)$$

The cumulative distribution function of LL random variable can be expressed in a regression form given as following way:

$$ln[(1-F(x))^{-1} - 1] = \alpha \ln(x) - \alpha \ln(\gamma) \quad (6)$$

For order statistics of sample and thus, regression model can be written as follows:

$$ln[1-(1-F(x_{i:n}))^{-1}] = \alpha \ln(x_{i:n}) - \alpha \ln(\gamma) \quad (7)$$

In the literature, the estimation of F(x) has been done with several formulas for both complete and censored samples (Zhang et al. 2006). For complete samples, Bernard's median rank estimator given in (8) is generally used.

$$\hat{F} = \frac{i-0.3}{n+0.4} \quad (8)$$

However, the form of estimator is changed for datasets containing censored values. For censored samples, the Herd–Johnson is one of the most used estimators for the estimate of F. Herd–Johnson method is non-parametric estimator for F(x) and is calculated based on only sample size and failure order number. This estimator is independent of the observation values of sample (Zhang et al. 2006). Its formula is given as follows:

$$\hat{R}_i = \frac{n-i+1}{n-i+2}\hat{R}_{i-1}$$
$$\hat{F}_i = 1-\hat{R}_i \quad (9)$$

3. Parameter estimators

In this section we have provided the classical and robust estimators of the parameters of LLD for the censored sample.

3.1. Classical estimators

In literature, ML, LS and moment estimators can be listed as classical estimators. However, moment estimators are not used due to constraints on parameters. Although ML estimation is widely-accepted method in censoring framework, generally LS is preferred due to ML' computational difficulties. It is known that the ML and LS estimators have no robustness property against contamination of outliers. In such cases, robust methods can be alternative estimation. In this section, we provide the description of robust estimators as well as ML and LS estimators.

3.1.1. The maximum likelihood estimator

The maximum likelihood estimator can be obtained using the following equation for censored data:

$$L(x) = \prod_{d\in D} f(x_d) \prod_{r\in R} S(x_r) \quad (10)$$

Where D and R denotes failure data set and censored data set, respectively. Using the probability density function (1) and survival function (4) and taking the logarithm the log-likelihood function is obtained as following:

$$\log L = m\log\alpha - m\log\gamma + (\alpha-1)\sum_{d\in D}\log\left(\frac{x_d}{\gamma}\right) - 2\sum_{d\in D}\log\left(1+\left(\frac{x_d}{\gamma}\right)^{\alpha}\right) + \sum_{r\in R}\log\left(\frac{\gamma^{\alpha}}{\gamma^{\alpha}+x_r^{\alpha}}\right) \quad (11)$$

Where m is the number of failure samples. After differentiating the log-likelihood function with respect to the parameters and equating to zero the following equations are obtained:

$$\alpha\sum_{d\in D}\log\left(\frac{x_d}{\gamma}\right) + 2\alpha\sum_{d\in D}\frac{\left(\frac{x_d}{\gamma}\right)^{\alpha}\log\left(\frac{x_d}{\gamma}\right)}{\left(\frac{x_d}{\gamma}\right)^{\alpha}+1} - \alpha\sum_{r\in R}\frac{x_r^{\alpha}\left(\log\gamma - \log x_r\right)}{\gamma^{\alpha}-x_r^{\alpha}} = m \quad (12)$$

$$m(1-\alpha) + 2\gamma\sum_{d\in D}\frac{\alpha\left(\frac{x_d}{\gamma}\right)^{\alpha}}{\gamma\left(\frac{x_d}{\gamma}\right)^{\alpha}+\gamma} + \gamma\sum_{r\in R}\frac{\alpha x_r^{\alpha}}{\gamma^{\alpha+1}+\gamma x_r^{\alpha}} = m \quad (13)$$

In this method, estimator may not have a closed form for most of the statistical distributions as well as LLD. Therefore numerical methods are used for obtaining estimates. It is known that ML estimation is asymptotically fully efficient for large sample sizes. As well as computational complexities of ML, it shows poor performance for small sample sizes. Furthermore, in the presence of censoring and contamination data cases, ML can be affected dramatically by contamination of outliers and thus lead to poor results (Arik 2014; Kantar and Yildirim, 2015; Basu et al. 2006).

3.1.2. The Least Squares Estimator

Since distribution function is expressed as a regression model as in (7), the parameters of LLD can be easily found with the help of LS estimation method.

One of the main advantages of using LS estimation for estimating parameters of statistical distributions is that its implementation is simple in the case of complete data, censoring data or data with outliers (Kantar, 2015). In this study, as in the Zhang (2006)' study, Herd-Johnson estimator is used when using LS method for censored data. However, it is known that LS is very sensitive to outliers and deviation from assumptions. (Arik and Kantar, 2014; Kantar and Yildirim, 2015; Kantar, 2016). This situation leads to switch to robust methods.

$$\ln\left[1-\left(1-F(x_{(i)})\right)^{-1}\right] = \alpha\ln(x_{(i)}) - \alpha\ln(\gamma) \quad (14)$$

Where $y_i = \ln\left[1-\left(1-F(x_{(i)})\right)^{-1}\right]$, $X_i = \ln(x_{(i)})$, $\beta_0 = -\alpha\ln(\gamma)$ and $\beta_1 = \alpha$. Thus, the model converts into linear model given as follows:

$$y = \beta_0 + \beta_1 X \quad (15)$$

Thus, LS estimates can be obtained for linear regression model (11).

3.2. Robust estimators

An outlier is an observation point that is distant from other observations because of the nature of the data or measurement errors. In the presence of censoring, it can be expressed that detection of contamination of outliers can be very difficult. In such cases, robust methods can be alternative. Robust estimators based on linear regression model have been well-accepted for censoring data due to easy estimation of F for censored data cases. In this estimation approach, the distribution function F(x) can be converted to the linear function of the parameters. Thus, parameter estimation problem is transformed into estimation of coefficients of linear regression model. In this study, robust estimators for linear regression model are used for distributional parameters of LLD for censored data. The considered methods are Tukey' M estimator, RM, LTS and LMS.

3.2.1. M-Estimator (Tukey)

Tukey, which is an M-estimator, is expressed as the following equation (16). It should also be noted that the ML estimator is a kind of M-estimator if the objective function can be taken as log likelihood of the sample (Maronna et al., 2006). Tukey's bisquare objective function is given as follows:

$$\rho(x) = \begin{cases} \left\{e\left[1-\left(\frac{e}{c}\right)^2\right]^2\right., & |x| \le c \\ 0, & |x| > c \end{cases} \quad (16)$$

By taking e as the studentized residual of the regression model, $\sum \rho'(e_i)$ is minimized with respect to the coefficients.

3.2.2. Least Median of Squares (LMS)

The estimator presented by Rousseeuw (1984) is based on the principle that minimizes the median squared error.

$$\min_{\beta} Med\{e_{(i)}^2\} \quad (17)$$

Zhang et al. (2006) have considered LMS method for the robust estimation of the parameters of the Weibull distribution. Arik (2014) also has evaluated the performance of LMS for the shape parameters of the Weibull and Pareto distributions in the case of samples with outliers (Kantar and Yildirim, 2015).

3.2.3. Least Trimmed Squares (LTS)

The coefficients of the regression model can be estimated by minimizing the sum of trimmed square error.

The objective function of this method is presented equation (18):

$$\min_{\beta} \sum_{i=1}^{h} \epsilon_{(i)}^2 \qquad (18)$$

Where $h = [n(1-\theta)+1]$ and θ is trimming proportion. Kantar and Yildirim (2015) have adapted the most of robust regression estimators for the Extended Burr Type III Distribution.

3.2.4. Repeated Median (RM)

Siegel's RM estimates of β_1 and β_0 for (15) are respectively presented equations (19) and (20):

$$\hat{\beta}_1 = Med_j Med_{i \neq j} \frac{y_j - y_i}{x_j - x_i} \qquad (19)$$

$$\hat{\beta}_0 = Med_j Med_{i \neq j} \frac{x_j y_i - x_i y_j}{x_j - x_i} \qquad (20)$$

(Siegel, 1982).

4. Simulation study

A simulations study has been implemented to assess the performance of ML, LS and robust estimators in the cases of censored data with/without outliers. Complete and censored samples were randomly generated from LL distribution with different shape and scale parameters. Censored samples were generated as in (Zhang et al., 2006; Wang, 2004). Procedures for generating censored samples are given as follows:

I) n random numbers, t_i, were generated from LLD ($i = 1,2,\ldots,n$).

II) n random numbers, u_i, were generated from Uniform (0,1) distribution ($i = 1,2,\ldots,n$).

III) If u_i > CL, where CL is censoring level, then the corresponding t_i is determined as a failure, if not, censored data.

IV) z_i were generated from Uniform (0,1) distribution ($i = 1,2,\ldots,k$), where k is selected as random number and also it is the number of censored data.

V) The values determined as censored data changed as $z_i t_i$.

VI) Finally combine the failures and censors.

We take sample size n as 10, 20 and 30, shape parameter α as 2, 4, 6 and 8, CL as 0.1 and 0.2. Without a loss of generality, scale parameter, γ, was taken to be equal to 1.

All the computations were done in MATLAB program. We use bias and mean square errors (RMSE) criteria to compare estimators. The biases and RMSEs for estimators were calculated based on 10.000 replications for different sample sizes.

$$Bias(\hat{\gamma}) = \left(\frac{1}{10000}\sum_{i=1}^{10000}\hat{\gamma}_i\right) - \gamma$$

$$Bias(\hat{\alpha}) = \left(\frac{1}{10000}\sum_{i=1}^{10000}\hat{\alpha}_i\right) - \alpha$$

$$RMSE(\hat{\gamma}) = \sqrt{\frac{1}{10000}\sum_{i=1}^{10000}(\hat{\gamma}-\gamma)^2}$$

$$RMSE(\hat{\alpha}) = \sqrt{\frac{1}{10000}\sum_{i=1}^{10000}(\hat{\alpha}-\alpha)^2}$$

The simulation results with different censored levels are given in Table 1-2 for different shape parameters and different sample sizes.

Table 1: RMSE and bias values of LLD for CL=0.1 (Without outliers).

n	10						
α	Est	ML	LS	Tukey	RM	LTS	LMS
2	Bias	-0.3185	0.3546	0.2481	-0.0866	0.1203	-0.0543
	RMSE	0.8257	0.6537	0.6849	0.9031	0.7943	1.2193
4	Bias	-0.6758	0.6953	0.4966	-0.2034	0.2342	-0.1192
	RMSE	1.7130	1.3165	1.3621	1.8573	1.6424	2.4256
6	Bias	-1.0205	1.0631	0.7485	-0.3182	0.3193	-0.2017
	RMSE	2.5517	1.9650	2.0472	2.8469	2.4960	3.7561
8	Bias	-1.3762	1.4115	1.0048	-0.3741	0.4814	-0.2118
	RMSE	3.3707	2.5952	2.6980	3.6150	3.1614	4.7082
n	20						
α	Est	ML	LS	Tukey	RM	LTS	LMS
2	Bias	-0.1457	0.2955	0.1972	0.0043	0.1139	0.0390
	RMSE	0.4902	0.4921	0.4704	0.5081	0.5081	0.6457
4	Bias	-0.3161	0.5903	0.4031	0.0144	0.2431	0.0785
	RMSE	0.9512	0.9503	0.9040	0.9762	0.9784	1.2846
6	Bias	-	0.9141	0.6319	0.0387	0.4058	0.1059

		0.4513					
	RMSE	1.4290	1.4392	1.3724	1.4790	1.4701	2.0339
8	Bias	-0.5878	1.2388	0.8338	0.0272	0.5224	0.1312
	RMSE	1.8689	1.9275	1.8170	1.9720	1.9734	2.6233
n	30						
α	Est	ML	LS	Tukey	RM	LTS	LMS
2	Bias	-0.0806	0.2702	0.1677	0.0313	0.1069	0.0519
	RMSE	0.3698	0.4197	0.3826	0.3918	0.4104	0.5122
4	Bias	-0.1851	0.5298	0.3297	0.0678	0.2324	0.1068
	RMSE	0.7337	0.8212	0.7451	0.7712	0.8034	1.0051
6	Bias	-0.3052	0.7895	0.5039	0.0842	0.3264	0.1705
	RMSE	1.1220	1.2270	1.1247	1.1649	1.2048	1.5005
8	Bias	-0.4112	1.0619	0.6578	0.0896	0.4237	0.1436
	RMSE	1.4719	1.6264	1.5049	1.5555	1.5790	2.1106

In view of the bias, the following conclusions can be made from Table 1 and Table 2 for CV=0.1 and CV=0.2, respectively that RM is the best estimator for almost all cases. Also LTS and LMS outperform ML, LS and Tukey. According to MSE criterion, while LS and Tukey provide similar and good performance for n=10 and 20, ML is the best for n= 30. Next to ML, Tukey shows good performance for n=30.

Table 2: RMSE and bias values of LLD for CL=0.2 (Without outliers).

n	10						
α	Est	ML	LS	Tukey	RM	LTS	LMS
2	Bias	-0.3511	0.3703	0.2931	-0.1488	0.1144	-0.1180
	RMSE	0.9173	0.7246	0.7595	1.1939	1.0035	1.5639
4	Bias	-0.7682	0.7385	0.5964	-0.4264	0.2307	-0.2392
	RMSE	2.0014	1.4828	1.5376	14.8223	2.0258	3.2978
6	Bias	-1.3004	1.0493	0.8326	-0.5393	0.2822	-0.3858
	RMSE	10.2250	6.5092	6.5350	7.6354	6.7942	7.3696
8	Bias	-1.6287	1.4722	1.1654	-0.6551	0.4342	-0.5191
	RMSE	4.0527	2.9754	3.0840	5.3276	4.1809	7.6809
n	20						
α	Est	ML	LS	Tukey	RM	LTS	LMS
2	Bias	-0.1425	0.3408	0.2484	0.0250	0.1585	0.0430
	RMSE	0.5088	0.5390	0.5173	0.5658	0.5667	0.7477
4	Bias	-0.3230	0.6996	0.5022	0.0648	0.3341	0.1366
	RMSE	1.0408	1.0844	1.0463	1.1283	1.1302	1.4765
6	Bias	-0.5123	1.0550	0.7429	0.1077	0.4800	0.2376
	RMSE	1.5847	1.6256	1.5558	1.6924	1.6979	2.1352
8	Bias	-0.6396	1.4426	1.0471	0.2316	0.7288	0.3670
	RMSE	2.0877	2.1704	2.0865	2.2091	2.2730	2.9430
n	30						
α	Est	ML	LS	Tukey	RM	LTS	LMS
2	Bias	-0.0632	0.3292	0.2354	0.0838	0.1712	0.1190
	RMSE	0.3685	0.4651	0.4297	0.4294	0.4577	0.5596
4	Bias	-0.1856	0.6422	0.4444	0.1451	0.3252	0.2031
	RMSE	0.7907	0.9387	0.8736	0.8840	0.9320	1.1225
6	Bias	-0.3114	0.9684	0.6806	0.2241	0.5190	0.3230
	RMSE	1.1720	1.3982	1.2937	1.2967	1.3635	1.6937
8	Bias	-0.4428	1.2581	0.8741	0.2828	0.6384	0.4071
	RMSE	1.5578	1.8354	1.6732	1.6777	1.7628	2.1908

Up to know, we only have researched the performance of the considered robust estimators of LLD in this paper in the case of censored data without outliers. Now, a new numerical study is performed for censored data with contamination of outliers. To generate outliers, we modified the 10% and %20 of the generated data by multiplying 10. The obtained results for % 10 contamination of outliers are given in Table 3 and 4 for CV=0.1 and CV=0.2, respectively.

Table 3: RMSE and bias values of LLD for CL=0.1,(10% contamination of outlies)

n	10						
α	Est	ML	LS	Tukey	RM	LTS	LMS
2	Bias	0.4610	1.1617	0.4788	0.0570	0.1600	-0.0623
	RMSE	0.5604	1.1701	0.8527	0.8988	0.8168	1.2961
4	Bias	1.6924	2.9840	0.9406	0.0962	0.2648	-0.2579
	RMSE	1.7293	2.9867	1.7962	1.8636	1.6349	3.0135
6	Bias	3.2539	4.9330	1.3834	0.1563	0.4009	-0.3863
	RMSE	3.2727	4.9344	2.6799	2.7459	2.4592	4.1267
8	Bias	4.9774	6.9150	1.8090	0.2424	0.5294	-0.4472
	RMSE	4.9891	6.9159	3.5372	3.5954	3.2709	5.2270
n	20						
α	Est	ML	LS	Tukey	RM	LTS	LMS
2	Bias	0.5314	1.0842	0.9969	0.1211	0.1383	0.0178
	RMSE	0.5723	1.0902	1.0300	0.5178	0.5055	0.7204
4	Bias	1.7584	2.8640	2.8730	0.2346	0.2719	0.0317
	RMSE	1.7755	2.8660	2.8758	1.0283	0.9934	1.4006
6	Bias	3.3046	4.7958	4.8177	0.3487	0.4009	0.0463
	RMSE	3.3134	4.7968	4.8188	1.5432	1.4862	2.0569
8	Bias	5.0206	6.7738	6.7993	0.4477	0.5195	0.0393
	RMSE	5.0263	6.7745	6.8001	2.1363	2.0744	2.9060
n	30						
α	Est	ML	LS	Tukey	RM	LTS	LMS
2	Bias	0.5552	1.0472	0.9807	0.1377	0.1254	0.0565
	RMSE	0.5805	1.0519	1.0055	0.4008	0.3998	0.5195
4	Bias	1.7745	2.8069	2.8258	0.2350	0.2213	0.0829
	RMSE	1.7856	2.8085	2.8276	0.7983	0.7876	1.0344
6	Bias	3.3245	4.7377	4.7682	0.3833	0.3456	0.1402
	RMSE	3.3303	4.7385	4.7691	1.2033	1.1866	1.5206
8	Bias	5.0440	6.7119	6.7476	0.5767	0.5338	0.2533
	RMSE	5.0477	6.7125	6.7482	1.6736	1.6223	2.1161

It can be deduced from Table 3 and Table 4 that LMS provide good performance for most of the considered sample cases in terms of bias criterion. Also RM, LTS and LMS outperform ML, LS and Tukey. According to RMSE, while RMSE values of RM, LTS and LMS decrease as sample sizes increase, those of non-robust increase. Also, RM, LTS and LMS outperform others in terms of RMSE.

For %20 of contamination, Table 5 and Table 6 are provided. For almost all considered sample and parameter cases except α=2 and n=10, RM, LTS and LMS show substantial improvement over ML, LS and Tukey both RMSE and bias criterion.

Table 4: RMSE and bias values of LLD for CL=0.2, (10% contamination of outlies)

n	10						
α	Est	ML	LS	Tukey	RM	LTS	LMS
2	Bias	0.5283	1.2198	0.7564	0.0926	0.2244	-0.0690
	RMSE	0.6163	1.2281	1.0423	1.0447	0.9314	1.5821
4	Bias	1.8277	3.0551	1.6259	0.2139	0.4122	-0.2468
	RMSE	1.8632	3.0585	2.2893	1.9997	1.7322	3.2026
6	Bias	3.4512	5.0103	2.4531	0.3359	0.5934	-0.3111
	RMSE	3.4713	5.0122	3.5818	2.9433	2.5361	4.5320
8	Bias	6.3772	7.1243	7.1340	1.7784	3.1664	-0.1191
	RMSE	6.3794	7.1250	7.1347	3.9574	5.0261	5.9631
n	20						
α	Est	ML	LS	Tukey	RM	LTS	LMS
2	Bias	0.5960	1.1428	1.1103	0.1911	0.2052	0.0795
	RMSE	0.6325	1.1487	1.1267	0.5619	0.5531	0.7408
4	Bias	1.8918	2.9441	2.9576	0.3934	0.4003	0.1475
	RMSE	1.9083	2.9466	2.9603	1.1201	1.0768	1.4245
6	Bias	3.4956	4.8848	4.9058	0.5698	0.5793	0.2049

	RMSE	3.5054	4.8863	4.9073	1.6535	1.6030	2.1193
8	Bias	5.2477	6.8622	6.8860	0.7238	0.7484	0.1901
	RMSE	5.2548	6.8633	6.8870	2.2509	2.1906	2.9704
n	30						
α	Est	ML	LS	Tukey	RM	LTS	LMS
2	Bias	0.6149	1.1085	1.0869	0.1928	0.1789	0.0953
	RMSE	0.6388	1.1132	1.0992	0.4536	0.4417	0.5670
4	Bias	1.9073	2.8935	2.9146	0.3941	0.3562	0.1954
	RMSE	1.9185	2.8954	2.9167	0.8905	0.8576	1.0799
6	Bias	3.5062	4.8301	4.8602	0.5459	0.4989	0.2270
	RMSE	3.5128	4.8314	4.8614	1.3405	1.2986	1.6717
8	Bias	4.3663	6.6581	6.6882	0.6059	0.7049	0.4118
	RMSE	4.3777	6.6594	6.6896	1.6622	1.6902	2.1244

Table 5: RMSE and bias values of LLD for CL=0.1, (20% contamination of outlies)

n	10						
α	Est	MLE	OLS	Tukey	RM	LTS	LMS
2	Bias	0.8714	1.2867	1.2918	0.3292	0.4944	0.0610
	RMSE	0.8875	1.2905	1.2957	0.8785	0.9818	1.3160
4	Bias	2.5159	3.1431	3.1519	0.6331	0.8516	-0.1425
	RMSE	2.5209	3.1443	3.1532	1.8774	1.8624	2.8700
6	Bias	4.3498	5.0907	5.1011	0.9892	1.1626	-0.2229
	RMSE	4.3523	5.0913	5.1018	2.5777	2.7007	3.7946
8	Bias	6.2630	7.0701	7.0815	1.3438	1.6278	-0.2983
	RMSE	6.2647	7.0705	7.0820	3.4348	3.7425	5.1763
n	20						
α	Est	MLE	OLS	Tukey	RM	LTS	LMS
2	Bias	0.9042	1.2158	1.2264	0.3160	0.3555	0.0555
	RMSE	0.9117	1.2185	1.2291	0.5734	0.6777	0.7005
4	Bias	2.5414	3.0463	3.0648	0.6345	0.6173	0.0888
	RMSE	2.5439	3.0471	3.0657	1.1369	1.2940	1.3419
6	Bias	4.3705	4.9877	5.0098	0.9090	0.9123	0.0815
	RMSE	4.3718	4.9882	5.0103	1.7371	1.9874	2.0860
8	Bias	6.2774	6.9633	6.9865	1.2808	1.3044	0.1369
	RMSE	6.2783	6.9637	6.9868	2.3241	2.6501	2.8037
n	30						
α	Est	MLE	OLS	Tukey	RM	LTS	LMS
2	Bias	0.9129	1.1834	1.1971	0.2740	0.2364	0.0462
	RMSE	0.9178	1.1854	1.1992	0.4667	0.5145	0.5232
4	Bias	2.5511	3.0055	3.0285	0.5691	0.4797	0.0958
	RMSE	2.5527	3.0062	3.0292	0.9380	1.0349	1.0413
6	Bias	4.3778	4.9425	4.9695	0.8603	0.7050	0.1834
	RMSE	4.3787	4.9429	4.9699	1.4460	1.5684	1.5322
8	Bias	6.2837	6.9173	6.9471	1.1467	1.0271	0.2151
	RMSE	6.2843	6.9176	6.9474	1.9168	2.1706	2.0957

Table 6: RMSE and bias values of LLD for CL=0.2, (20% contamination of outlies)

n	10						
α	Est	MLE	OLS	Tukey	RM	LTS	LMS
2	Bias	0.9174	1.3287	1.3335	0.4228	0.7502	0.1571
	RMSE	0.9323	1.3328	1.3376	1.0360	1.1766	1.6655
4	Bias	2.5999	3.1964	3.2041	0.8570	1.5024	0.0190
	RMSE	2.6050	3.1980	3.2057	2.0469	2.4086	3.5693
6	Bias	4.4597	5.1471	5.1559	1.3601	2.3711	0.0226
	RMSE	4.4626	5.1481	5.1569	3.2331	3.7236	4.3087
8	Bias	6.3763	7.1251	7.1350	1.7863	3.1304	-0.1577
	RMSE	6.3784	7.1258	7.1357	3.8911	5.0325	5.9691
n	20						
α	Est	MLE	OLS	Tukey	RM	LTS	LMS
2	Bias	0.9483	1.2569	1.2677	0.4093	0.6844	0.1120
	RMSE	0.9555	1.2597	1.2706	0.6446	0.9603	0.7668
4	Bias	2.6263	3.1049	3.1217	0.8282	1.5099	0.1351
	RMSE	2.6289	3.1060	3.1229	1.3024	2.1487	1.5180

6	Bias	4.4784	5.0500	5.0694	1.2516	2.4565	0.2028
	RMSE	4.4800	5.0507	5.0701	1.9580	3.4836	2.2312
8	Bias	6.3943	7.0243	7.0449	1.6066	3.2728	0.1918
	RMSE	6.3954	7.0248	7.0454	2.6476	4.7291	3.0746
n	30						
α	Est	MLE	OLS	Tukey	RM	LTS	LMS
2	Bias	0.9612	1.2287	1.2431	0.3698	0.6636	0.0936
	RMSE	0.9658	1.2308	1.2452	0.5449	0.9102	0.5748
4	Bias	2.6349	3.0657	3.0875	0.7569	1.5475	0.1737
	RMSE	2.6366	3.0666	3.0885	1.1082	2.1347	1.1184
6	Bias	4.4878	5.0108	5.0367	1.1355	2.4739	0.2576
	RMSE	4.4888	5.0114	5.0373	1.6535	3.4390	1.7001
8	Bias	5.8196	6.8916	6.9246	0.9789	0.7456	0.3711
	RMSE	5,8220	6,8922	6,9253	1,8925	1,7619	2,2056

5. Conclusions

The log-logistic regression model, based on LLD, is one of the alternative parametric models in the case of samples containing censored values. In this study, we have studied the robust estimators of LLD in the cases of censored data with/without outliers. The classical and robust estimators of the parameters of LLD have been evaluated by means of Monte Carlo simulation study. The simulation results for the shape parameter of LLD can be summarized as follows: robust alternatives, RM, LMS and LTS are able to improve over ML and LS for the censored sample cases with outliers. For the cases of sample without outliers, the most of the considered robust estimators for LLD can show better performance than the others according to bias criterion.

References

Arik, I. Robust Estimation Methods For Estimating Distributional Parameters. Master of science thesis, Anadolu University, Eskisehir, TURKEY, 2014.

Bennett S. Log-logistic regression models for survival data. Appl. Stat., 1983, 32 (2), pp.165-171.

Byers, R. H., Morgan, W. M., Darrow, W. W., Doll, L., Jaffe, H. W., Rutherford G. and Hessol N. Estimating AIDS infection rates in the San Francisco cohort. AIDS, 1988, 2: 207-210.

Chen, Z. Estimating the shape parameter of the log-logistic distribution. Int. jour. rel., qual. and safety eng., 2006, 13, pp. 257-266.

Chen, Z. M. Exact confidence interval for the shape parameter of a log-logistic distribution. Journal of statistical computation and simulation, 1997, 56(3), pp.193-211.

Cox, D. R. and Oakes, D. Analysis of survival data. 1984, Chapman and Hall, London.

Gupta, R. D, Akman, O. and Lvin, S. A. Study of Log-Logistic model in survival analysis. Biometrical journal, 1999, 41: 431-443.

Guure, C. B. Inference on the Log-logistic model with right censored data. Austin biom. and biostat., 2015, 2(1): 10-15.

Kantam, R. R. L, Rao, A. V. and Rao, G. S. Linear estimation in scaled Log-logistic distribution. International journal of agricultural and statistical sciences, 2008, 4, 1, 11-22.

Kantar, Y. M. and Arik, I. M-estimation of log-logistic distribution parameters with outliers. International journal of agricultural and statistical sciences, 2014, vol. 10, no. 1, 7-16.

Kantar, Y. M. and Yildirim, V. Robust estimation for parameters of the Extended Burr Type III distribution. Communications in statistics - simulation and computation, 2015, 44(7), 1901-1930.

Kantar, Y. M. Generalized least squares and weighted least squares estimation methods for distributional parameters. 2015, Revstat statistical journal, in press.

Karasoy, D. and Ata N. Stratified Cox regression model for non-proportional hazards and an application on breast cancer patients, Turkiye Klinikleri Tip Bilimleri Dergisi, 2008, 28(3), 327-332.

Klein, J. P. and Moeschberger, M. L. Survival analysis: techniques for censored and truncated data. 1997, Springer-Verlag, New York.

Kleinbaum, D. G. Survival analysis: a self learning text. 1996, Springer, New York.

Lawless, J. F. Statistical models and methods for lifetime data analysis. 1982, Wiley and Sons, New York.

Lee, E. T. and Wang, J. W. Statistical methods for survival data analysis. 2003, Wiley and Sons, New York.

Maronna, R. A., Martin, D. R. and Yohai, V. J. Robust statistics theory & methods. 2006, Wiley, England.

Marubini, E. and Valsecchi, M. G. Analysing survival data from clinical trials and observational Studies. 1995, John Wiley & Sons, Chichester, UK.

May, M., Royston, P., Egger, M., Justice, A. C. and Sterne J. A. C. Development and validation of a prognostic model for survival time data: application to prognosis of HIV positive patients treated with antiretroviral therapy. Statistics in medicine, 2004, 23 (15), 2375-2398.

Nelson, W. B. Applied life data analysis. 2004, John Wiley & Sons, New Jersey, USA.

Pourhoseingholi, M. A., Hajizadeh, E., Dehkordi, B. M., Safaee, A., Abadi, A. and Zali, M. Z. Comparing cox regression and parametric models for survival of patients with gastric carcinoma. Asian pacific journal of cancer prevention, 2007, 8, 412-416.

Rousseeuw, P. J. Least median of squares regression. Journal of American statistical association, 1984, 79, pp. 871-880.

Siegel, A. F. Robust regression using repeated medians. Biometrika, 1982, 69, 242-244.

Singh, A. and Nocerino, J. Robust estimation of mean and variance using environmental data sets with below detection limit observations. Chemometrics and intelligent lab. sys., 2002, 60: 69-86.

Singh, V. P. and Guo, H. Parameter estimation for 2-parameter Log-logistic distribution (LLD2) by maximum entropy, Civil. eng. sys., 1995, 12(4), 343-357.

Wang, W. Refined rank regression method with censors. Qual. reliab. eng. int., 2004, 20: 667–678. doi: 10.1002/qre.575.

Wei, L. J. The accelerated failure time model: A useful alternative to the cox regression model in survival analysis. Statistics in medicine, 1992, 11 (14–15): 1871–1879.

Zhang, L. F., Xie, M. and Tang, L. C. Robust regression using probability plots for estimating the Weibull shape parameter. Quality and reliability engineering international, 2006, 22, 905-917